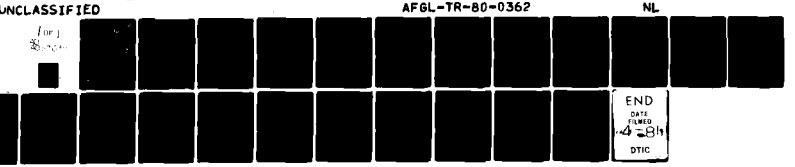
AD-A097 039 UNIVERSITY OF CENTRAL FLORIDA ORLANDO DEPT OF MATHEM--ETC F/G 12/1
LEAST SQUARES FITTING OF DISTRIBUTIONS USING NON-LINEAR REGRESS--ETC(U)
SEP 80 M HEUSER, P N SOMERVILLE, S J BEAN F19628-80-C-0004

UNCLASSIFIED AFGL-TR-80-0362 NL



AFGL-TR- 80 - 0362

LEAST SQUARES FITTING OF DISTRIBUTIONS USING NON-LINEAR REGRESSION

Mark Heuser
Paul N. Somerville
Steven J. Bean

Department of Mathematics and Statistics
University of Central Florida
Orlando, Florida 32816

Scientific Report, No. 1

30 September 1980

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: AFGL-TR-80-0362	2. GOVT ACCESSION NO.: AD-A097039
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): LEAST SQUARES FITTING OF DISTRIBUTIONS USING NON-LINEAR REGRESSION	5. TYPE OF REPORT & PERIOD COVERED: Scientific Report, No. 1, 1 Oct 79 - 30 Sep 80
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Mark Heuser, Paul N. Somerville, Steven J. Bean	8. CONTRACT OR GRANT NUMBER(s): F19628-80-C-0004
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Central Florida, Department of Mathematics and Statistics, P.O. Box 25000 Orlando, Florida 32816	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62101F, 6677009 AF
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Geophysics Laboratory, Hanscom AFB, MA 01731, Contract Monitor: I. I. Gringorten, LYD	12. REPORT DATE: 30 Sep 80
	13. NUMBER OF PAGES: 28
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release, Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Non-Linear Regression
Mathematical Models
Cumulative Distribution
Linearization
Curve Fitting
Least Squares
Climatology

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report is concerned with techniques of estimating the parameter values of mathematical models. It begins with a brief discussion of mathematical models and how they relate to the physical situations they attempt to describe. Next, the method of least squares as it applies to both linear and nonlinear models is investigated. Special emphasis is given to nonlinear models with the linearization method. Finally, the report is concluded with an example from climatology.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ACKNOWLEDGEMENT

This research was motivated and partially supported by Contract Nos. F19628-77-R-0120 and F19628-80-C-0004 with Cambridge Research Laboratories of the U. S. Air Force and monitored by I. Gringorten.

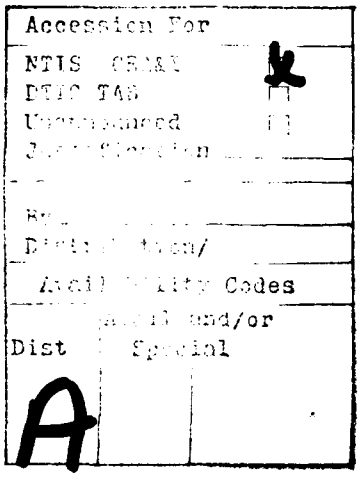


TABLE OF CONTENTS

I. INTRODUCTION

This report is concerned with techniques of estimating the parameter values of mathematical models. It begins with a brief discussion of mathematical models and how they relate to the physical situations they attempt to describe. Next, the method of least squares as it applies to both linear and nonlinear models is investigated. Special emphasis is given to nonlinear models with the linearization method. Finally, the report is concluded with an example from climatology.

2. MATHEMATICAL MODELS

A problem that arises in many fields is that of determining what relationships exist between variables. Suppose, for example that an experiment has produced a sample of data consisting of a number of simultaneous observations of a group of variables. It would be convenient to put the data in the form of an array as shown below.

$$\begin{array}{cccccc} y_1 & x_{11} & x_{12} & \cdots & x_{1k} \\ y_2 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ y_i & x_{i1} & x_{i2} & & x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ y_n & x_{n1} & x_{n2} & & x_{nk} \end{array}$$

Each row represents one simultaneous observation of all the variables, and each column represents all the observed values of a particular variable. One variable (which one depending on the nature of the experiment) is designated the response or dependent variable (denoted by y), and the others are called independent variables (denoted by $x_1, x_2, \ldots, x_k$). From the sample of data, a researcher would like to make inferences about possible relationships existing between the independent and dependent variables. Mathematical models are commonly used to describe and measure the strength of these relationships. They also provide a means of predicting the value of the dependent variable for any given setting of the independent variables.

The relationship between the dependent and independent variables is rarely a functional one, where the behavior of the dependent variable can be predicted exactly from the independent variables. Experiments are of course vulnerable to error, and the results may include effects from outside sources. Consequently, the mathematical models under consideration must allow for noise in the data. Models of this kind are called probabilistic models and have the form

$$y = f(\underline{x},\underline{\theta}) + r,$$

where $\underline{x} = (x_1, x_2, \ldots, x_k)$ is a vector of independent variables,

$\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ is a vector of parameters, and r is the residual

or difference between the observed value y and the predicted value $f(\underline{x}, \theta)$ of the dependent variable.

The residual r is made up of two components. The pure experimental error (ε) and the error due to lack of fit (e). The pure experimental error may be due to inaccuracies in the instruments, human mistakes, etc.. It is an inherent part of the data at this point and beyond our control. Some assumptions are usually made about ε: that it is random in nature and has expectation zero. If this is so, then the data has no systematic source of error that would tend to skew the results in one direction or another. The error due to lack of fit is the error in the model, and it too may arise from a number of sources. Perhaps not all the significant variables are accounted for, or perhaps the relationship expressed by the model is not an appropriate one to use.

With these considerations, the model may be rewritten

$$y = f(\underline{x}, \underline{\theta}) + e + \varepsilon \quad \text{where} \quad r = e + \varepsilon.$$

The function f and the computed values of the parameters are ideally those that minimize e and average out ε.

Once a particular function f is chosen for the model, the next problem is to compute the parameter values. The parameter values are those that best describe the particular sample of data available, and are used as estimates of the parameter values that best fit the larger population. The methods of computing the parameter values depend on the for of f; specifically, whether f is linear or nonlinear with respect to its parameters. Linear models have the form

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k + r,$$

and models not of this form, with respect to the parameters, are nonlinear. Some examples of linear and nonlinear models are shown below.

Some Linear Models

$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + r$ (cubic polynomial, one independent variable)

$y = \theta_0 \sin x_1 + \theta_2 x_2 + \theta_3 x_2 \sin x + r$ (two independent variables).

Some Nonlinear Models

$y = 1 - \exp\{-\theta_1 x^{\theta_2}\} + r$ (used in modeling visibility, see reference [5])

$y = 1 - (1 - x^{\theta_1})^{\theta_2} + r$ (used in modeling skycover, see reference [6]).

3. METHOD OF LEAST SQUARES IN THE LINEAR CASE

Suppose that the set of data

$$\begin{array}{cccc} y_1 & x_{11} & x_{12} & \cdots x_{1k} \\ y_2 & x_{21} & x_{22} & \cdots x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ y_i & y_{i1} & x_{i2} & \cdots x_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ y_n & y_{n1} & x_{n2} & x_{nk} \end{array}$$

is given and a linear model, $Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_k x_k + r$, is chosen to express a relationship between the dependent variable y and the independent variables $x_1, x_2, \ldots, x_k$.

Regardless of what values may be arbitrarily assigned to the parameters $\theta_0, \theta_1, \ldots, \theta_k$, for each observation in the data set, there exists an r_i so that

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_k x_{ik} + r_i \quad .$$

This equation is called an observation equation. Solving for r_i gives

$$r_i = y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \ldots - \theta_k x_{ik} \quad .$$

Intuitively, one would like to find the values for $\theta_0, \theta_1, \ldots, \theta_k$ that minimize $|r_i|$ for each observation in the data set. The method of least squares finds those parameter values that minimize instead the sum of the squared residuals over all the observations. That is, those parameter values that minimize

$$RSS = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \ldots - \theta_k x_{ik})^2$$

are found.

Observe that RSS is a function of the parameters $\theta_0, \theta_1, \ldots, \theta_k$; y_i, x_{i1}, x_{i2}, ..., x_{ik} all being known values (the data). To minimize RSS, its partial derivative with respect to each θ_j is taken and set equal to zero. The resulting system of $k + 1$ linear equations in $k + 1$ unknowns (often called the normal equations) has as its solution the vector of parameter values that minimize RSS.

The partial derivatives are shown below:

$$\frac{\partial RSS}{\partial \theta_0} = -2 \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \ldots - \theta_k x_{ik})$$

$$\frac{\partial RSS}{\partial \theta_1} = -2 \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \ldots - \theta_k x_{ik}) x_{i1}$$

$$\vdots$$

$$\frac{\partial RSS}{\partial \theta_j} = -2 \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \ldots - \theta_k x_{ik}) x_{ij}$$

$$\vdots$$

$$\frac{\partial RSS}{\partial \theta_k} = -2 \sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_{i1} - \theta_2 x_{i2} - \ldots - \theta_k x_{ik}) x_{ik}$$

Setting them equal to zero and simplifying them algebraically yields the system of normal equations:

$$\theta_0 n + \theta_1 \Sigma x_{i1} + \theta_2 \Sigma x_{i2} \ldots \theta_k \Sigma x_{ik} = \Sigma y_i$$

$$\theta_0 \Sigma x_{i1} + \theta_1 \Sigma x_{i1}^2 + \theta_2 \Sigma x_{i2} x_{i1} + \ldots + \theta_k \Sigma x_{ik} x_{i1} = \Sigma y_i x_{i1}$$

$$\vdots$$

$$\theta_0 \Sigma x_{ij} + \theta_1 \Sigma x_{i1} x_{ij} + \theta_2 \Sigma x_{i2} x_{ij} + \ldots + \theta_k \Sigma x_{ik} x_{ij} = \Sigma y_i x_{ij}$$

$$\vdots$$

$$\theta_0 \Sigma x_{ik} + \theta_1 \Sigma x_{i1} x_{ik} + \theta_2 \Sigma x_{i2} x_{ik} + \ldots + \theta_k \Sigma x_{ik}^2 = \Sigma y_i x_{ik} \; .$$

The equations may be further simplified by using MATRIX notation. Let $\underline{Y}$ be the column vector of values of the dependent variable:

$$\underline{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}$$

, and let $\underline{X}$ be the matrix of values of the independent variables, appended on theleft with a column vector of 1's:

$$\underline{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}$$

. Lastly, define $\underline{\theta}$ to be the column vector of parameters in the model and $\underline{r}$ to be the column vector of residuals:

$$\underline{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_j \\ \vdots \\ \theta_k \end{bmatrix} \qquad \underline{r} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_i \\ \vdots \\ r_n \end{bmatrix}$$

. Then the system of n observation equations

$$\begin{aligned}
y_1 &= \theta_0 + \theta_1 x_{11} + \theta_2 x_{12} + \ldots + \theta_k x_{1k} + r_1 \\
y_2 &= \theta_0 + \theta_1 x_{21} + \theta_2 x_{22} + \ldots + \theta_k x_{2k} + r_2 \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
y_i &= \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \ldots + \theta_k x_{ik} + r_i \\
&\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
y_n &= \theta_0 + \theta_1 x_{n1} + \theta_2 x_{n2} + \ldots + \theta_k x_{nk} + r_n
\end{aligned}$$

may be written simply as

$$\underline{Y} = \underline{X}\,\underline{\theta} + \underline{r} \quad .$$

Furthermore, the normal equations may be rewritten in terms of matrix multiplication:

$$\begin{bmatrix} n & \Sigma x_{i1} & \Sigma x_{i2} & \cdots & \Sigma x_{ik} \\ \Sigma x_{i1} & \Sigma x_{i1}^2 & \Sigma x_{i2}x_{i1} & \cdots & \Sigma x_{i1}x_{ik} \\ \Sigma x_{i2} & \Sigma x_{i1}x_{i2} & \Sigma x_{i2}^2 & \cdots & \Sigma x_{i2}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \Sigma x_{ij} & \Sigma x_{i1}x_{ij} & \Sigma x_{i2}x_{ij} & \cdots & \Sigma x_{ij}x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ \Sigma x_{ik} & \Sigma x_{i1}x_{ik} & \Sigma x_{i2}x_{ik} & \cdots & \Sigma x_{ik}^2 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_j \\ \vdots \\ \theta_k \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma y_i x_{i1} \\ \Sigma y_i x_{i2} \\ \vdots \\ \Sigma y_i x_{ij} \\ \vdots \\ \Sigma y_i x_{ik} \end{bmatrix}$$

Noting that the matrix on the left is $\underline{X}^T\underline{X}$ and the matrix on the right is $\underline{X}^T\underline{Y}$, the entire system of normal equations may be expressed by the single matrix equation

$$(\underline{X}^T\ \underline{X})\underline{\theta} = (\underline{X}^T\ \underline{Y}) \quad .$$

Finally, solving for $\underline{\theta}$ gives

$$\hat{\underline{\theta}} = (\underline{X}^T\ \underline{X})^{-1}\ (\underline{X}^T\ \underline{Y}) \quad .$$

This is a very powerful result, since it holds for any linear model and corresponding set of $n \geq k + 1$ observations (where n is the number of observations and $k + 1$ is the number of parameters; that is, there should be at least as many observations in the data set as parameters in the model). The first step would be to construct the matrices $\underline{X}$ and $\underline{Y}$ from the data. Then the parameter values that minimize RSS are those that satisfy the linear system

$$(\underline{X}^T\ \underline{X})\ \underline{\theta} = (\underline{X}^T\ \underline{Y}) \quad .$$

Most computer installations have routines for solving systems of linear equations. As a warning, however, this particular system is often ill-conditioned, so round-off error is an important factor in the computations.

4. METHOD OF LEAST SQUARES IN THE NONLINEAR CASE

Suppose that the sample of data

$$
\begin{array}{llll}
y_1 & x_{11} & x_{12} & \cdots x_{1k} \\
y_2 & x_{21} & x_{22} & \cdots x_{2k} \\
\vdots & & & \vdots \\
y_i & x_{i1} & x_{i2} & \cdots x_{ik} \\
\vdots & & & \vdots \\
y_n & x_{n1} & x_{n2} & \cdots x_{nk}
\end{array}
$$

is given and a nonlinear model, $y = f(\underline{x},\underline{\theta}) + r$, is chosen to express a relationship between the dependent variable y and the independent variables $x_1, x_2, \ldots, x_k$.

Proceeding in much the same way as for the linear model, regardless of what values may be arbitrarily assigned to the parameters, for each observation in the data set there exists an r_i so that

$$y_i = f(\underline{x}_i, \underline{\theta}) + r_i \quad .$$

Solving for r_i gives

$$r_i = y_i - f(\underline{x}_i,\underline{\theta}).$$

The method of least squares minimizes the quantity

$$RSS = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - f(x_i,\theta))^2$$

with respect to its parameters by setting the partial derivatives equal to zero and solving the resulting system of equations. The partial derivatives are of the form shown below:

$$\frac{\partial RSS}{\partial \theta_j} = -2 \sum_{i=1}^{n} (y_i - f(\underline{x}_i,\underline{\theta})) \frac{\partial f(\underline{x}_i,\underline{\theta})}{\partial \theta j} \qquad \text{for } j = 1, \ldots, p.$$

Therefore, the system of normal equations is

$$\sum_{i=1}^{n} (y_i - f(\underline{x}_i, \underline{\theta})) \frac{\partial f(\underline{x}_i, \underline{\theta})}{\partial \theta_j} = 0 \quad \text{for } j = 1, \ldots, p \quad .$$

This approach may be fine for some nonlinear models, but for others difficulties arise. The model may be so complex that the partial derivatives are difficult to obtain. Furthermore, since the model is nonlinear the partial derivatives and normal equations are also nonlinear. Systems of nonlinear equations can be difficult to solve, and iterative techniques are almost always required. Also, more than one solution to the system may exist, corresponding to the critical values of RSS. In this case, each solution must be tested to determine which produces the minimum value of RSS. For these reasons, other methods have been devised for estimating the parameter values of a nonlinear model.

5. THE LINEARIZATION METHOD

The linearization method of nonlinear regression has three distinguishing features. First, it is an iterative procedure. Given an initial estimate of the parameter values, a "better" estimate is computed with each successive iteration. The parameter values are refined in this manner an unknown number of times until finally some stopping criterion is met. Second, it is not self-starting. This means that the method does not provide a way of obtaining that initial estimate of the parameter values. The researcher must use whatever information is at hand to make the first guess. Lastly, linear approximations of the nonlinear model are used to compute the parameter values, so the techniques already developed for linear models are applicable.

Let $\underline{T} = (t_1, t_2, \ldots, t_p)$ be the current best estimate of $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$. To "linearize" the nonlinear model

$$y = f(\underline{x}, \underline{\theta}) + r,$$

a first order Taylor series expansion of f about $\underline{T}$ is made. The expansion is a linear polynomial in $(\theta_j - t_j)$ of the form

$$P(\underline{x}, \underline{\theta}) = f(\underline{x}, \underline{T}) + \sum_{j=1}^{p} \frac{\partial f(\underline{x}, \underline{T})}{\partial \theta_j} (\theta_j - t_j) \quad .$$

For θ close to $\underline{T}$, $P(\underline{x}, \underline{\theta}) \simeq f(\underline{x}, \underline{\theta})$. Therefore, substituting P for f gives a linear approximation of the nonlinear model.

$$y = P(\underline{x}, \underline{\theta}) + r$$

$$y = f(\underline{x}, \underline{T}) + \sum_{j=1}^{p} \frac{\partial f(\underline{x}, \underline{T})}{\partial \theta_j} (\theta_j - t_j) + r$$

$$y - f(\underline{x}, \underline{T}) = \frac{\partial f(\underline{x}, \underline{T})}{\partial \theta_1}(\theta_1 - t_1) + \frac{\partial f(\underline{x}, \underline{T})}{\partial \theta_2} (\theta_2 - t_2) + \ldots + \frac{\partial f(\underline{x}, \underline{T})}{\partial \theta_p} (\theta_p - t_p) + r \quad .$$

In terms of the linear approximation, the system of observation equations is

$$y_i - f(\underline{x}_i, \underline{T}) = \frac{\partial f(\underline{x}_i, \underline{T})}{\partial \theta_1} (\theta_1 - t_1) + \frac{\partial f(\underline{x}_i, \underline{T})}{\partial \theta_2} (\theta_2 - t_2) + \ldots + \frac{\partial f(\underline{x}_i, \underline{T})}{\partial \theta_p} (\theta_p - t_p) + r_i$$

for $i = 1, 2, \ldots, n$. Note that y_i, $\underline{x}_i$, $\underline{T}$, $f(\underline{x}_i, \underline{T})$, and $\frac{\partial f(\underline{x}_i, \underline{T})}{\partial \theta_j}$ are all known values or values that can be computed directly. The only unknowns are the θ_j in $(\theta_j - t_j)$. Letting

$W_{ij} = \frac{\partial f(\underline{x}_i, \underline{T})}{\partial \theta_j}$, $Z_i = y_i - f(\underline{x}_i, \underline{T})$, $\delta_j = \theta_j - t_j$, and the matrices

$$\underline{W} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1p} \\ W_{21} & W_{22} & \cdots & W_{2p} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ W_{n1} & W_{n2} & \cdots & W_{np} \end{bmatrix}, \quad \underline{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ \cdot \\ Z_n \end{bmatrix} = \begin{bmatrix} y_1 - f(x_1, T) \\ y_2 - f(x_2, T) \\ \cdot \\ \cdot \\ \cdot \\ y_n - f(x_n, T) \end{bmatrix}, \quad \underline{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \cdot \\ \delta_p \end{bmatrix} = \begin{bmatrix} \theta_1 - t_1 \\ \theta_2 - t_2 \\ \cdot \\ \cdot \\ \cdot \\ \theta_p - t_p \end{bmatrix},$$

$$\underline{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \cdot \\ \cdot \\ \cdot \\ \theta_p \end{bmatrix}, \quad \underline{T} = \begin{bmatrix} t_1 \\ t_2 \\ \cdot \\ \cdot \\ \cdot \\ t_p \end{bmatrix}, \quad \text{and } \underline{r} = \begin{bmatrix} r_1 \\ r_2 \\ \cdot \\ \cdot \\ \cdot \\ r_n \end{bmatrix},$$

then the system of observation equations may be written

$$\underline{Z} = \underline{W} \ \underline{\delta} + \underline{r} \ .$$

Applying linear least squares theory, the solution for $\underline{\delta}$ that minimizes

$$RSS_{lin} = \sum_{i=1}^{n} \left[y_i - f(x_i, T) - \sum_{j=1}^{p} \frac{\partial f(x_i, T)}{\partial \theta_j} (\theta_j - t_j) \right]^2$$

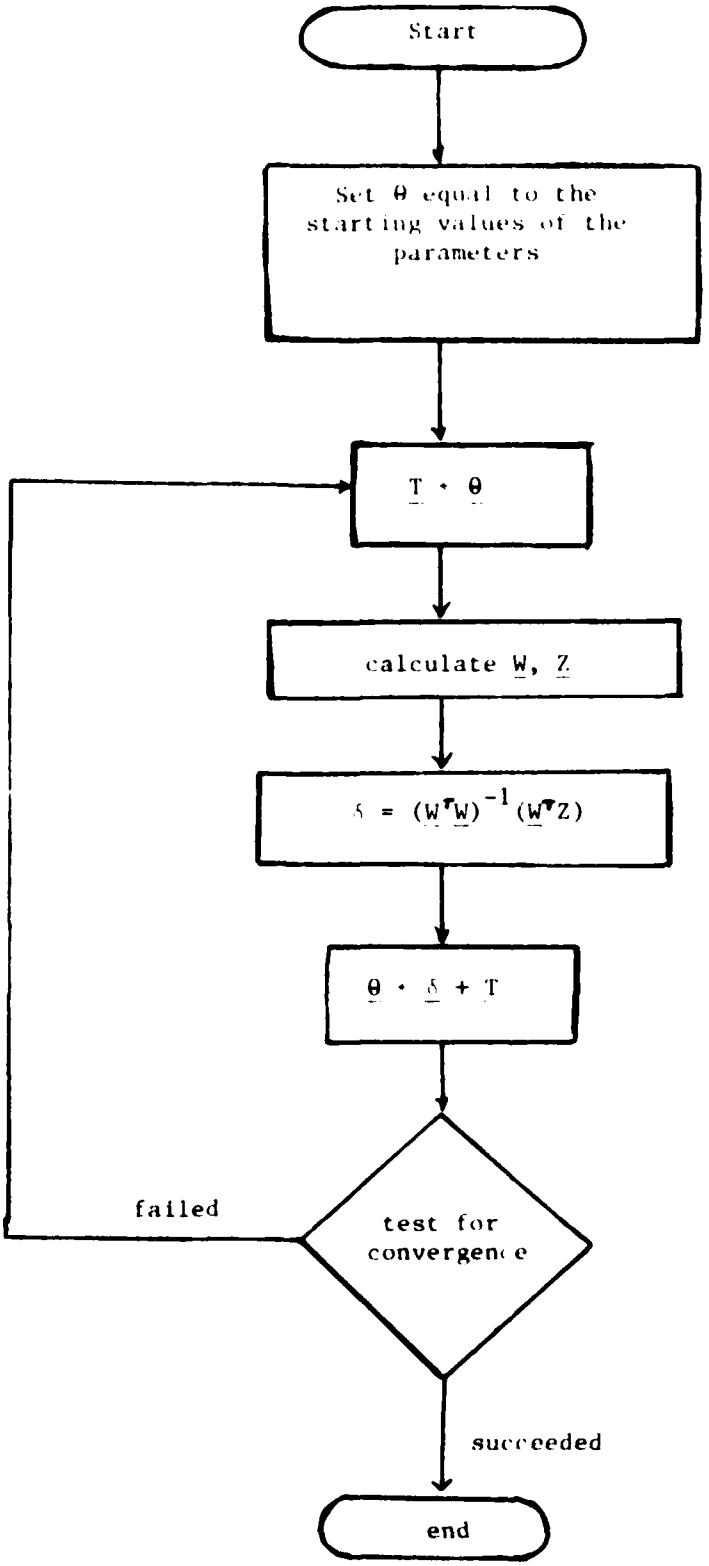
is given by

$$\underline{\delta} = (\underline{W}^T W)^{-1} \quad (\underline{W}^T \underline{Z}) \ ,$$

and the new estimate of $\underline{\theta}$ is

$$\underline{\theta} = \underline{\delta} + \underline{T} \ .$$

$\underline{\delta}$ may be thought of as being a correction vector, so that when added to the old estimate of the parameter values, a newer, "better" estimate is obtained. The method now calls for substituting the new estimate of the parameter values in for T, and repeating the procedure. The following flowchart outlines the algorithm.


Start
Set θ equal to the starting values of the parameters
T ← θ
calculate W, Z
δ = (W^T W)^-1 (W^T Z)
θ ← δ + T
test for convergence
failed
succeeded
end

Obviously some test must be made to determine when to halt the procedure, as indicated in the flowchart by the test for convergence. Any one of several tests may be used. Intuitively one would halt the procedure when the estimates of the parameter values approach a limiting value; i.e., when the difference between two consecutive estimates is small. If A is any matrix, let |A| denote the largest absolute value of the elements of A. Then one possible test for convergence is the relation

$$\left| \frac{\Theta}{T} \right| < c_1$$

where c_1 is some prescribed constant value and the division is an element-wise division between the matrices Θ and T. This relation tests whether the largest proportional change in the estimates is less than c_1. It is most appropriate when the relative magnitudes of the parameters are unknown. Typical values for c_1 are in the range 10^{-5} to 10^{-6}.

Another test for convergence might be performed by comparing the sum of the squared residuals for two consecutive estimates of the parameters. The sum of the squared residuals of the nonlinear model, given by

$$RSS(\Theta) = \sum_{i=1}^{n} y_i - f(x_i, \Theta)$$

is a function of Θ and is a measure of the goodness of fit of the model. Recall that in the algorithm the matrices Θ and T represent two consecutive estimates of the parameter values. Hence, convergence may be tested by the relation.

$$\left| \frac{RSS(\Theta) - RSS(T)}{RSS(T)} \right| < c_2$$

where c_2 is some constant. This is appealing in the sense that it directly tests for a difference in the fit between the two models. It is conceivable that a wide range of values for the parameters might give approximately the same fit, in which case this test would assume convergence before the other. However, if the only purpose in estimating the parameter values is to obtain a good fit, then the test is appropriate.

Lastly, the importance of good starting values should be stressed. Convergence in the linearization method is not guaranteed; its success, failure and speed may well depend on how good the starting values are

(among other things). (To allow for the possibility that convergence fails, a bound on the number of iterations allowed should be included in the algorithm.) As mentioned earlier, the researcher is left up to his/her own ingenuity in finding starting values, making use of whatever information is available. One suggestion is to plot a grid of the RSS computed at various parameter values to get an idea of where minimums might occur. For difficult problems, more complicated variations of the linearization method may be employed. Also, the NLIN procedure of the Statistical Analysis System (SAS) is a software package that performs an extensive analysis of the nonlinear regression problem.

6. AN EXAMPLE FROM CLIMATOLOGY

Climatological data is collected on a daily basis at several locations around the world. The data includes observations of various surface weather conditions such as rainfall, skycover, ceiling, and visibility. The visibility observations recorded over the years are grouped so that for a given location, month, and hour period there are 15 data points of the form (x,y), where x is distance in statute miles and y is the proportion of the time in the past that visibility has been less than or equal to x. The following data comes from Goose, Newfoundland in February during the hours of 0900 to 1100:

x	y
.25	.003
.3125	.015
.5	.017
.625	.023
.75	.025
1.00	.038
1.25	.065
1.5	.073
2.00	.095
2.5	.126
3	.134
4	.159
5	.188
6	.209
10	.261

Problem: For a given location, month, and hour period, build a probability model relating distance (x) to the proportion of the time that visibility has been less than or equal to x (y).

The Weibull cumulative frequency distribution function has been used to model the visibility data for several of these locations. In the model

$$y = f(\underline{x},\underline{\theta}) + r = 1-e^{-\theta_1 x^{\theta_2}} + r,$$

the vector $\underline{x}$ of independent variables consists only of x, and $\underline{\theta}$, the vector of parameters, is (θ_1,θ_2). The model is obviously nonlinear in its parameters, so the linearization method is used to compute the parameter values.

The first problem is that of finding starting values for the parameters. These may be obtained by taking two observations from the

data set, substituting them into the formula

$$y = 1 - e^{-\theta_1 x^{\theta_2}} ,$$

and solving for θ_1 and θ_2, which is equivalent to fitting the curve through the two data points exactly. If (x_i,y_i) and (x_j,y_j) are the two observations chosen, then we have the system

$$y_i = 1 - e^{-\theta_1 x_i^{\theta_2}}$$

$$y_j = 1 - e^{-\theta_1 x_j^{\theta_2}}$$

which, after some manipulation, yields

$$\theta_2 = \frac{\ln\left[\frac{\ln(1-y_i)}{\ln(1-y_j)}\right]}{\left[\ln\frac{x_i}{x_j}\right]} \qquad \text{and} \qquad \theta_1 = \frac{-\ln\left[1-y_i\right]}{x_i^{\theta_2}}$$

Using (1.00, 0.038) and (5.00, 0.188) from the data shown earlier, the starting values are

$$\hat{\theta}_1 = .0387408 \qquad \text{and} \qquad \hat{\theta}_2 = 1.045$$

The next problem is to "linearize" the model. Recall that if $T = (t_1, t_2, \ldots, t_p)$ is the current best estimate of $\underline{\theta} = (\theta_1,\theta_2,\ldots,\theta_p)$, then the nonlinear model is approximated by

$$y - f(x,T) = \frac{\partial f(\underline{x},\underline{T})}{\partial\theta_1}(\theta_1-t_1) + \frac{\partial f(\underline{x},\underline{T})}{\partial\theta_2}(\theta_2-t_2) + \ldots + \frac{\partial f(\underline{x},\underline{T})}{\partial\theta_p}(\theta_p-t_p) + r .$$

$f(x,\underline{\theta}) = 1 - e^{-\theta_1 x^{\theta_2}}$, and the partial derivatives are

$$\frac{\partial f(x,T)}{\partial\theta_1} = x^{t_2} \cdot e^{-t_1 x^{t_2}} \qquad \text{and} \qquad \frac{\partial f(x,T)}{\partial\theta_2} = t_1 \cdot \ln(x) \cdot x^{t_2} \cdot e^{-t_1 x^{t_2}} ,$$

so the linear approximation is

$$y - \left[1-e^{-t_1 x^{t_2}}\right] = \left[x^{t_2} \cdot e^{-t_1 x^{t_2}}\right](\theta_1-t_1) + \left[t_1 \cdot \ln(x) \cdot x^{t_2} \cdot e^{-t_1 x^{t_2}}\right](\theta_2-t_2) + r .$$

From this the matrices $\underline{Z}$ and $\underline{W}$ may be identified:

$$\underline{Z} = \begin{bmatrix} y_1 - [1 - e^{-t_1 x_1^{t_2}}] \\ y_2 - [1 - e^{-t_1 x_2^{t_2}}] \\ \vdots \\ y_n - [1 - e^{-t_1 x_n^{t_2}}] \end{bmatrix}, \qquad \underline{W} = \begin{bmatrix} x_1^{t_2} \cdot e^{-t_1 x_1^{t_2}} & t_1 \cdot \ell n(x_1) \cdot x_1^{t_2} \cdot e^{-t_1 x_1^{t_2}} \\ x_2^{t_2} \cdot e^{-t_1 x_2^{t_2}} & t_1 \cdot \ell n(x_2) \cdot x_2^{t_2} \cdot e^{-t_1 x_2^{t_2}} \\ \vdots & \\ x_n^{t_2} \cdot e^{-t_1 x_n^{t_2}} & t_1 \cdot \ell n(x_n) \cdot x_n^{t_2} \cdot e^{-t_1 x_n^{t_2}} \end{bmatrix}$$

Now that starting values have been found and we know how to compute $\underline{Z}$ and $\underline{W}$, we merely follow the algorithm expressed in the flowchart. The visibility data from Goose, Newfoundland converged after 7 iterations. Convergence was assumed when the proportional change in the RSS for two consecutive estimates was less than 1×10^{-8}. Below are the intermediate and final estimates, computed by a program written in PROC MATRIX of the Statistical Analysis System. Double precision arithmetic was used.

Iteration	θ_1	θ_2	RSS
0 (Starting values)	.0387408	1.045	.010229
1	.0560605	.734919	.00336779
2	.0546557	.787792	.00261866
3	.0553048	.77971	.00261354
4	.055203	.780875	.00261344
5	.055218	.780708	.00261344
6	.0552159	.780732	.00261344
7	.0552162	.780729	.00261344

References

[1] Draper, N. R., and H. Smith, Applied Regression Analysis, New York: John Wiley & Sons, Inc. 1966.

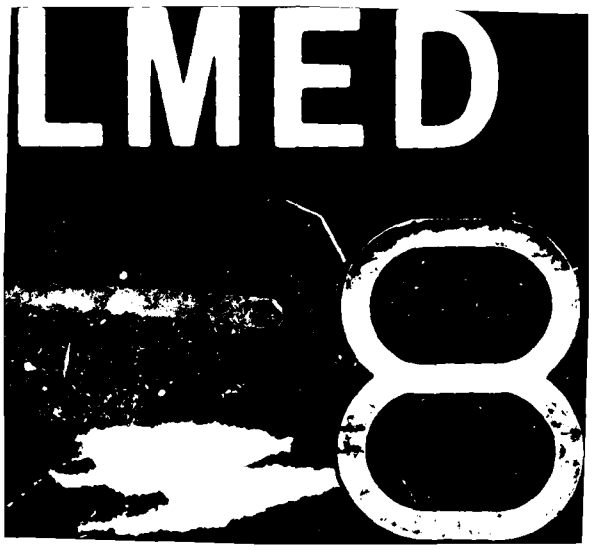
[2] Helwig, Jane T. and Kathryn A. Council, eds. SAS User's Guide 1979 Edition, Raleigh, North Carolina: SAS Institute, Inc. 1979.

[3] Hornbeck, Robert W., Numerical Methods, New York: Quantum Publishers, Inc. 1975

[4] Ott, Lyman, An Introduction to Statistical Methods and Data Analysis, North Scituate, Massachusetts: Duxbury Press, 1977.

[5] Somerville, Paul N. and Steven J. Bean, A New Model for Sky Cover, Air Force Geophysics Laboratory - Technical Report 79-0219.

[6] Somerville, Paul N. Steven J. Bean and Sherrill Falls, Some Models for Visibility, Air Force Geophysics Laboratory - Technical Report 79-0144.


LMED
8